Ubiquitous late competence genes in Bacillus species indicate the presence of functional DNA uptake machineries

Kovacs, Akos T.; Smits, Wiep Klaas; Mironczuk, Aleksandra M.; Kuipers, Oscar P.; Mirończuk, Aleksandra M.

Published in:
Environmental Microbiology

DOI:
10.1111/j.1462-2920.2009.01937.x

Publication date:
2009

Environmental Microbiology (2009) **11**(8), 1911–1922 doi:10.1111/j.1462-2920.2009.01937.x

Minireview

Ubiquitous late competence genes in *Bacillus* species indicate the presence of functional DNA uptake machineries

Ákos T. Kovács,[1] Wiep Klaas Smits,[1†] Aleksandra M. Mirończuk[1] and Oscar P. Kuipers[1,2*]
[1]*Department of Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9751 NN Haren, the Netherlands.*
[2]*Kluyver Centre for Genomics of Industrial Fermentation, Delft, the Netherlands.*

Summary

Natural competence for genetic transformation, i.e. the ability to take up DNA and stably integrate it in the genome, has so far only been observed in the bacterial kingdom (both in Gram-negative and Gram-positive species) and may contribute to survival under adverse growth conditions. *Bacillus subtilis*, the model organism for the *Bacillus* genus, possesses a well-characterized competence machinery. Phylogenetic analysis of several genome sequences of different *Bacillus* species reveals the presence of many, but not all genes potentially involved in competence and its regulation. The recent demonstration of functional DNA uptake by *B. cereus* supports the significance of our genome analyses and shows that the ability for functional DNA uptake might be widespread among *Bacilli*.

Introduction

The Gram-positive bacterium *Bacillus subtilis* is a soil-dwelling member of the genus *Bacillus*, which comprises commercially interesting (for instance *B. amyloliquefaciens*, *B. licheniformis* and *B. stearothermophilus*) as well as pathogenic species (such as *B. cereus* and *B. anthracis*). It can be isolated from many environments (Vilain *et al*., 2006) and it is regarded a model organism for Gram-positives, because of its long history of scientific research and good experimental amenability (Sonenshein, 2002). In addition to this, *B. subtilis* was one of the first bacteria for which cellular differentiation was recognized (spore formation and competence development). As spore formation is accompanied by easily visible morphological changes, much of the early research on cellular differentiation was aimed at the isolation, mapping and classification of sporulation genes (Hoch, 1971; Piggot and Coote, 1976). Groundbreaking work of Anagnostopoulos and Spizizen in the early 1960s established a competence regime for *B. subtilis* (Spizizen, 1958; Anagnostopoulos and Spizizen, 1961), making it genetically accessible and paving the way for advanced molecular research. As a result, the characterization of competence and competence genes has become one of the longest standing fields of investigation for this bacterium. Since that time, many genes and proteins involved in competence have been characterized (for reviews see Hamoen *et al*., 2003; Chen and Dubnau, 2004).

Received 3 February 2009; accepted 13 March 2009. *For correspondence. E-mail o.p.kuipers@rug.nl; Tel. (+31) 50 3632093; Fax (+31) 50 3632348. †Present address: Department of Biology, Building 68–540D, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

DNA uptake apparatus in *Bacillus subtilis*

The ability to take up naked DNA has been detected in eubacteria, including Gram-positive (e.g. *Bacillus*, *Streptococcus*) and Gram-negative species (e.g. *Campylobacter*, *Haemophilus*, *Helicobacter*, *Neisseria*, *Vibrio*) (Lorenz and Wackernagel, 1994). It may help bacteria to survive under adverse conditions such as nutrient limitation (Finkel and Kolter, 2001; Claverys *et al*., 2006; Palchevskiy and Finkel, 2006). In addition, it has been suggested that the thermophilic traits of the highly transformable bacterium *Thermus thermophilus*, which allow it to survive under extreme temperatures, were acquired via gene transfer (Averhoff, 2009). The DNA uptake machinery in most competent bacteria is related to Type II secretion systems and Type IV pili (Dubnau, 1999). In the case of Gram-positive organisms, such as *B. subtilis*, the presence of a thick peptidoglycan layer necessitates the modulation of the cell wall make-up in order for the DNA

to access a membrane-located receptor, $ComE_A$, which contains a DNA binding domain in its C-terminus (Provvedi and Dubnau, 1999). The proteins encoded by the *comG* operon resemble those of type IV pili (Dubnau, 1997) and are thought to form a structure called the competence pseudopilus (Chen and Dubnau, 2004; Chen *et al*., 2006). The $ComG_A$ protein is believed to be a traffic NTPase, providing the energy for the assembly of the competence pseudopilus. Single molecule measurements of DNA translocation into *B. subtilis* have demonstrated that DNA uptake depends on the transmembrane proton motif force (Maier *et al*., 2004). The $ComG_A$ protein localizes to the cell-pole during competence, and hence functional transport is believed to take place at that site (Hahn *et al*., 2005). $ComG_B$ is a polytopic membrane protein that is required for pilus assembly. The major pseudopilin is $ComG_C$, while $ComG_D$, $ComG_E$ and $ComG_G$ are minor pseudopilins. All of these proteins are produced as prepilins, and require cleavage of an N-terminal sequence by the prepilin peptidase ComC before assembly into a pseudopilus is possible (Chung and Dubnau, 1995; Chen *et al*., 2006). Recently it was reported that intramolecular disulfide bonds in the major pseudopilin $ComG_C$, dependent on the thiol-disulfide oxidoreductase pair BdbCD, are required to stabilize the protein (Meima *et al*., 2002; Chen *et al*., 2006). Strikingly, the competence pseudopilus is dispensable for DNA binding in the absence of a cell wall (Provvedi and Dubnau, 1999), supporting the hypothesis that it modulates the cell wall in such a way that the DNA gains access to the $ComE_A$ receptor protein. The receptor subsequently delivers the DNA to the permease, $ComE_C$. Oligomers of the $ComE_C$ protein, that is essential for DNA uptake (Hahn *et al*., 1987; Inamine and Dubnau, 1995; Dubnau, 1999), form an aqueous pore through which the DNA is transported into the cell (Draskovic and Dubnau, 2005). Like $ComG_C$, the protein contains an intramolecular disulfide bond that is probably introduced by BdbCD (Draskovic and Dubnau, 2005). $ComF_A$ is a membrane-associated protein (Londono-Vallejo and Dubnau, 1994) that is structurally similar to an ATP-dependent family of helicases and may act as the motor protein for DNA transport (Londono-Vallejo and Dubnau, 1993). Alternatively, its helicase activity may be required for unwinding of incoming double-stranded DNA. However, only single-stranded DNA (ssDNA) was shown to be taken up into the cells (Venema *et al*., 1965; Dubnau and Cirigliano, 1972). The non-transforming strand is degraded by an unidentified nuclease, which is presumably located on the outside of the membrane (Chen and Dubnau, 2004), and the degradation products are released into the medium (Dubnau and Cirigliano, 1972). The importance of DNA processing for transport is reinforced by the notion that the introduction of double strand breaks (cleavage) is required for efficient uptake, facilitated by the competence-induced nuclease NucA (van Sinderen *et al*., 1995; Provvedi *et al*., 2001). In order to reconstitute replicative plasmids or allow recombination of the transforming ssDNA with the host chromosome, the DNA has to be shielded from the action of nucleases by association with a competence-induced ssDNA binding protein (Eisenstadt *et al*., 1975), which likely corresponds to DprA (Smf) (Mortier Barriere *et al*., 2007; Tadesse and Graumann, 2007), SsbB (YwpH) (Lindner *et al*., 2004; Hahn *et al*., 2005; Morrison *et al*., 2007) and/or RecA (Lovett *et al*., 1989; Kidane and Graumann, 2005). Indeed, colocalization of these proteins with $ComG_A$ and $ComF_A$ *in vivo* has been observed at the cell pole using fluorescence resonance energy transfer experiments (Kramer *et al*., 2007). Plasmids are reconstituted and stably maintained, while ssDNA can be integrated into a homologous double-stranded DNA duplex mediated by the RecA ATPase. Its function depends in part on a helicase/nuclease complex that is formed by the AddAB proteins in *B. subtilis* (Haijema *et al*., 1995; Arnold and Kowalczykowski, 2000; Kidane and Graumann, 2005), although it has been noted that the *addAB* genes were not identified as ComK-dependent in a DNA array analysis of the ComK regulon (Hamoen *et al*., 2002). CoiA (YjbF) contributes to the recombination process in a way that is not fully understood (Desai and Morrison, 2006; 2007). Notably, an extensive network of interactions exists between proteins that act after DNA uptake (Kidane and Graumann, 2005; Kramer *et al*., 2007; Mortier Barriere *et al*., 2007).

ComK, the master regulator for competence development in *Bacillus*

The transcription of the genes that encode the DNA-binding, -uptake and -recombination proteins is controlled by the auto-activating competence transcription factor ComK (van Sinderen and Venema, 1994). In *B. subtilis*, the induction of this protein is strictly regulated at the level of *comK* transcription as well as post-translationally. Transcription of *comK* is repressed by binding of AbrB, CodY and Rok to the *comK* promoter region (Hamoen *et al*., 2003), whereas the ComK protein is trapped by MecA and targeted for proteolytic degradation by ClpCP (Turgay *et al*., 1998). High cell density is a prerequisite for optimal competence development, as at least two independent quorum sensing pathways induce the production of anti-adaptor protein ComS, which liberates ComK from the proteolytic complex (D'Souza *et al*., 1994; Hamoen *et al*., 1995; Turgay *et al*., 1997; Prepiak and Dubnau, 2007).

Through the use of whole genome DNA microarrays, it was established that ComK is directly or indirectly responsible for the activation of ~100 genes (Berka

et al., 2002; Hamoen *et al*., 2002; Ogura *et al*., 2002). ComK activates transcription by binding through the minor groove to specific sequences, so called K-boxes that are composed of two AT-boxes with the consensus sequence AAAA-N5-TTTT. The boxes are separated by a spacing of a discrete number of helical turns, which places them on the same side of the DNA-helix (Hamoen *et al*., 1998).

Distribution of genes encoding the competence machinery in various *Bacilli*

Competence development within the *Bacillus* genus has so far been described for *B. subtilis*, *B. licheniformis* and *B. amyloliquefaciens* (Spizizen, 1958; Thorne and Stull, 1966; Koumoutsi *et al*., 2004), but not for *B. cereus*, *B. anthracis* or other *Bacillus* species. With the availability of the genome sequence of other members of *Bacillaceae* (Takami *et al*., 2000; Ivanova *et al*., 2003; Read *et al*., 2003; Rasko *et al*., 2004; Rey *et al*., 2004; Han *et al*., 2006), our attention was drawn to the high number of homologues of competence-related genes in other species.

Up to now, the only functional competence machinery that has been characterized in detail is the one of *B. subtilis*. Similarly, the genetic network controlling the differentiation into the competent state has only been thoroughly characterized for this species. To gain insight in the function and evolution of the DNA uptake machinery, we have analysed all the completely sequenced bacterial genomes of *Bacillaceae* and other closely related bacteria, one *Oceanobacillus* and two *Geobacillus* species, as they are deposited in the NCBI database (20 genomes in January 2009). These BLAST analyses revealed the presence of many orthologous genes putatively involved in DNA uptake, and are visualized with Genesis software (Sturn *et al*., 2002) in Fig. 1.

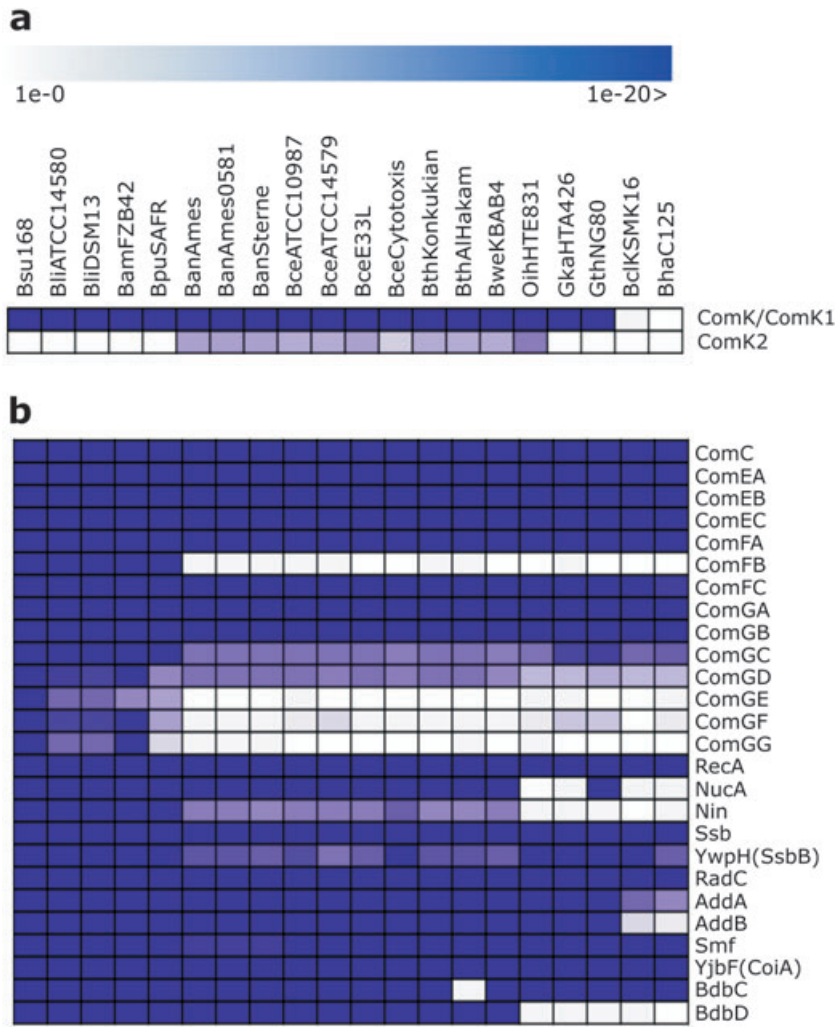

Fig. 1. Presence of homologues of competence regulator proteins (A) and competence-related structural proteins (B) in *Bacillus* and closely related species. Results of BLAST searches were visualized with Genesis 1.6 software: white is absent (with *e*-value of E-0), dark blue is present (*e*-value < E-20). BLAST analysis was performed with *B. subtilis* protein sequences against translated protein database of a given genome. Where no hit (or bellow E-05) was found additional TBLASTN was performed. Protein names are indicated on the right. Bsu, *B. subtilis*; Bam, *B. amyloliquefaciens*; Bli, *B. licheniformis*; Bpu, *B. pumilus*; Ban, *B. anthracis*; Bce, *B. cereus*; Bth, *B. thuringiensis*; Bwe, *B. weihenstephanensis*; Oih, *O. iheyensis*; Gka, *G. kaustophilus*; Gth, *G. thermodenitrificans*; Bcl, *B. clausii*; Bha, *B. halodurans*).

Presence of the master regulator, ComK, in members of the *Bacillus* genus

In *B. subtilis*, competence development is strictly dependent on the ComK protein. Homologues of the *comK* gene were present in most species, but could not be detected in *B. clausii* KSMK16 and *B. halodurans* C-125. The presence or absence of *comK* in these species does not seem to be correlated to the presence of other competence genes. Assuming a similar dependency for competence development on ComK, this suggests that the gene may have been lost recently. In support of this, some other genes are either absent (*nucA*, *nin*, *addB*) or less conserved (*comGD*) compared with species that do harbour a *comK* gene (Fig. 1). In addition, the *comG* operon of *B. clausii* and *B. halodurans* contains an inserted tRNA region. Both these species are alkaliphilic and it is possible that diversification into an alkaline environment is not compatible with natural competence. Disruption of competence gene homologues by tRNA genes is reminiscent of the prophage insertion in *comK* of *Listeria monocytogenes* (Borezee *et al*., 2000).

Interestingly, species from the *B. cereus* group of *Bacilli* (*B. cereus, B. anthracis, B. thuringiensis* and *B. weihenstephanensis*) contain two genes with homology to ComK (hereafter referred to as ComK1 and ComK2) (Fig. 2B). The putative ComK proteins show 61–62% (ComK1) and 44–48% (ComK2) similarity to *B. subtilis* ComK respectively. The ComK1 proteins, in general, are similar in length to the *B. subtilis* ComK protein, whereas the ComK2 proteins of the *B. cereus* group appear to be C-terminally truncated by 22–32 amino acids (Fig. 3).

Strikingly, the ComK/ComK1 proteins of outliers *G. kaustophilus*, *G. thermodenitrificans* and *O. iheyensis* are similarly 15, 10 and 23 amino acids shorter than the *B. subtilis* ComK protein respectively, raising the possibility that such a specific truncation may be relevant for the function of the protein. *Oceanobacillus iheyensis* ComK2 is 39 amino acids shorter. Previous studies on *B. subtilis* ComK have shown that a 25–35-amino-acid C-terminal truncation is incapable of transcriptional activation of a specific late competence promoter (P*comG*) (Susanna *et al*., 2006), although the protein retained its ability to bind to DNA. Potentially, the multimerization of the ComK protein, presumably required for *comG* activation in *B. subtilis*, is affected. It will be of great interest to determine the oligomeric state of the various ComK1 and ComK2 proteins. Notably, both ComK1 and ComK2 of *B. cereus* ATCC14579 possess DNA binding activity (A.M. Mirończuk, Á.T. Kovács and O.P. Kuipers, unpublished).

The C-terminal part of the *B. subtilis* ComK protein shows homology to the DNA binding domain to two High

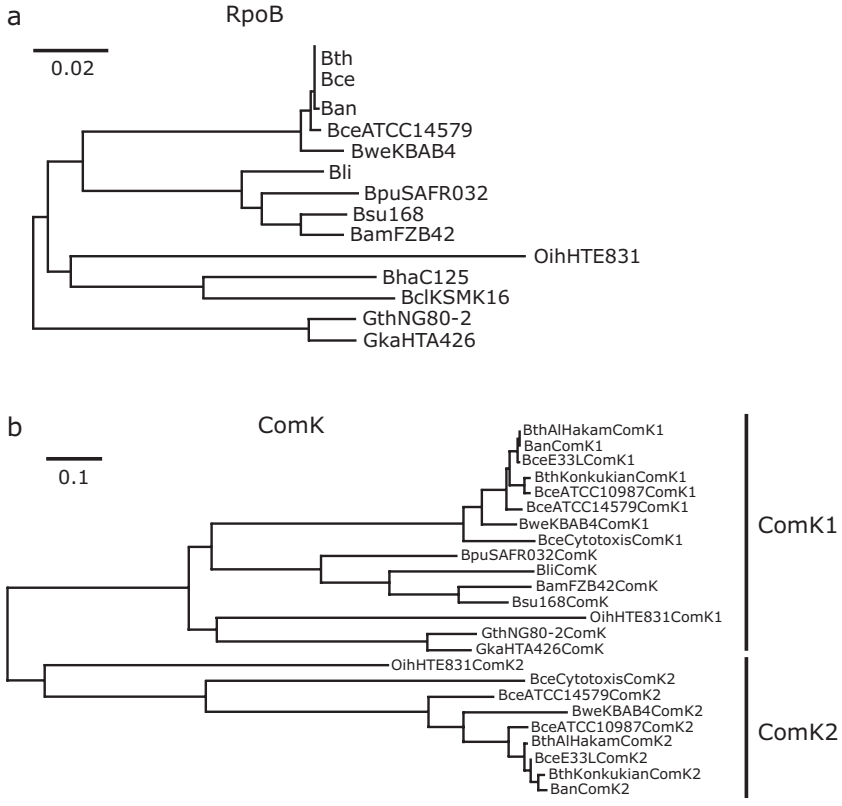

Fig. 2. Phylogenetic trees based on the RpoB (A) and ComK (B) protein sequences. Proteins were aligned using Clustal W (Thompson *et al*., 1994) and an evolutionary tree was generated using Treecon software (Van de Peer and De Wachter, 1994). The grouping of ComK/ComK1 and ComK2 proteins is indicated. For abbreviations of species names see Fig. 1. In the case of *B. anthracis* and *B. licheniformis* strains the proteins are grouped and strain names are not indicated as the corresponding protein sequences are identical.

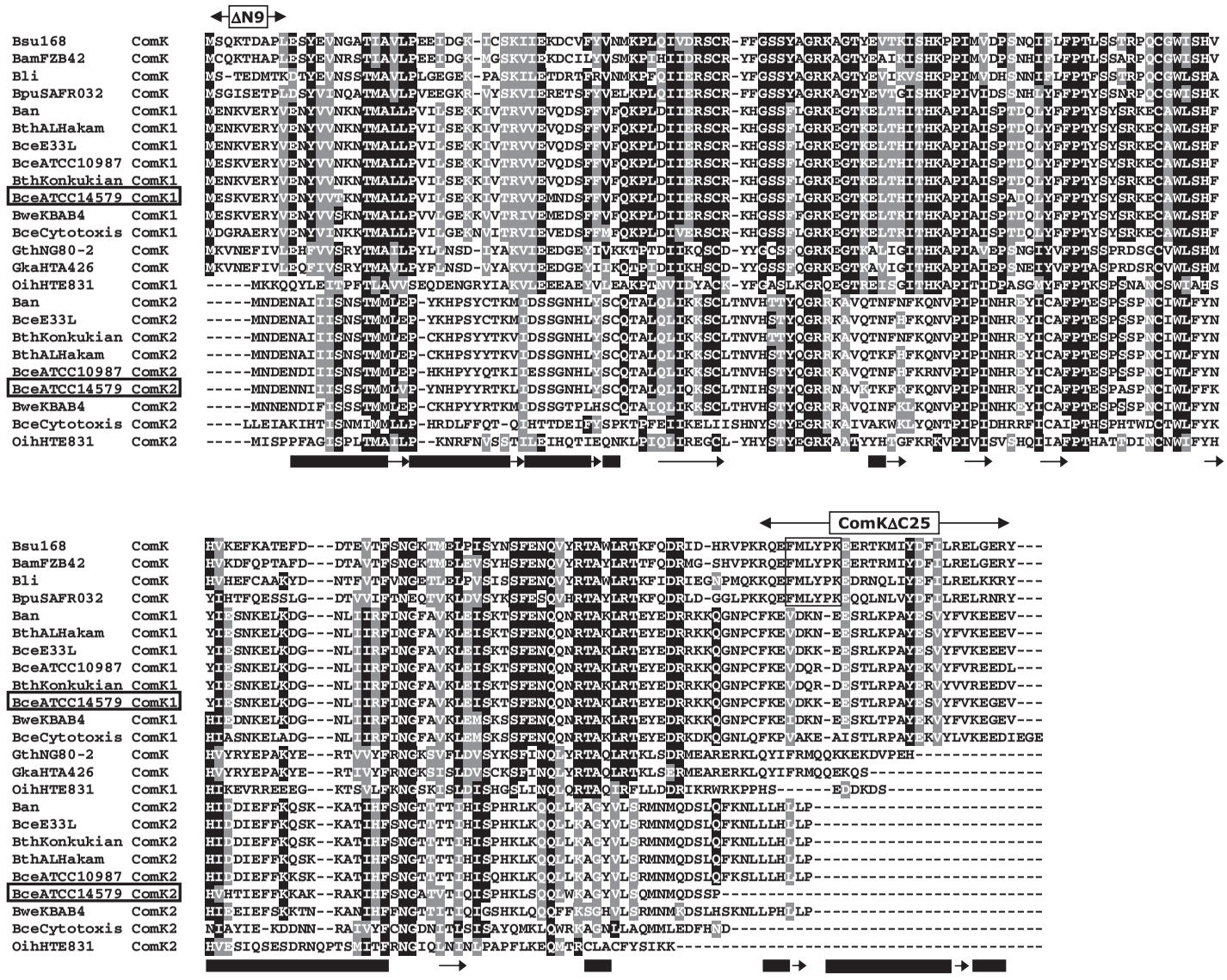

Fig. 3. Multiple alignment of ComK homologues. The ComK proteins of *B. cereus* ATCC14579 are highlighted with boxes. Black background represents conserved amino acids and grey background represents similar amino acids. Alignment was performed using Clustal W (Thompson *et al.*, 1994), and presented using the Boxshade 3.21 program. The N- and C-terminal deletions analysed by Susanna and colleagues (2006) are marked. Boxed amino acid residues indicate the residues involved in interaction with MecA (Prepiak and Dubnau, 2007). Alpha-helices and beta-sheets of *B. subtilis* ComK protein are indicated with rectangles and arrows under the alignment respectively. For abbreviations of species names see Fig. 1. In the case of *B. anthracis* and *B. licheniformis* strain names are not indicated as ComK protein sequences are identical within the different strains of the same species.

Mobility Group proteins, SRY and TCF-1, that like ComK bind through the minor grove of the DNA and induce bending. However, as both C-terminally truncated *B. subtilis* ComK (Susanna *et al*., 2006) and ComK2 of *B. cereus* – that lacks this region – have DNA binding activity, it seems unlikely that the region of homology is the sole DNA binding domain of ComK. It is conceivable, however, that this region is required for modulating the DNA topology, a property that may be critical for transcriptional activation by *B. subtilis* ComK (Smits *et al*., 2007).

A quest for ComK binding sites in the upstream region of putative competence related genes in *B. cereus* ATCC14579 reveals the presence of AT-boxes, although spacing between these sites mostly does not correspond to the spacing in functional K-boxes found in *B. subtilis*. This makes it likely that the *B. cereus* ComK proteins recognize a different sequence or that they bind as a dimer to AT-boxes, rather than as tetramer to K-boxes.

The regulation of ComK levels

Recently, it was found that the C-terminus of ComK contains a region that is required and sufficient for the interaction with MecA (Prepiak and Dubnau, 2007). This FMLYPK motif can be found in the more closely related *B. subtilis*, *B. amyloliquefaciens*, *B. licheniformis* and *B. pumilus*, but not in any other ComK homologues (Fig. 3 and Prepiak and Dubnau, 2007). This indicates either (i) that a different interaction site is responsible for controlling ComK levels in these species or (ii) that it does not involve a MecA homologue. In either case, the regulatory events upstream of *comK* expression may differ significantly from that of *B. subtilis*. One possibility is that the regulation does not involve quorum sensing-dependent production of an anti-adaptor protein, such as ComS. In support of this, *comS* could not be identified in the $srfA_A$ genes of the *B. cereus* group. Moreover, genes from the quorum sensing pathway could not unambiguously be identified in most species (Fig. S1). It has to be noted that the prediction of small open reading frames (ORFs) is not trivial and the *comS* gene may be located in other regions of the chromosome.

Not surprisingly considering their function in other pathways (Hamoen *et al*., 2003), the pleiotropic regulators that directly or indirectly control *comK* transcription (e.g. DegU, CodY, AbrB and Spo0A) were almost universally identified in all species (see Fig. S1). As noted before, Rok may have been acquired recently, as it is only present in the *B. subtilis/amiloliquefaciens/pumilus/licheniformis* group (Albano *et al*., 2005). Whether these regulators have a role similar to the one in *B. subtilis* competence development remains to be established.

Structural genes for DNA uptake machinery in various *Bacillus* species

For all species analysed, homologues for most of the structural genes encoding DNA uptake machinery could be identified. Two striking exceptions are the $comF_B$ and $comG_{EFG}$ genes.

The $comF_B$ gene was present only in *B. subtilis* and its close relatives *B. amyloliquefaciens*, *B. licheniformis* and *B. pumilus*, whereas the downstream $comF_C$ gene was present in all species analysed. The first gene of the *comF*-operon encodes a helicase-like protein essential for transformation. The function of $ComF_B$ is unknown, although transformation frequencies are slightly reduced in a $comF_B$ mutant (Londono-Vallejo and Dubnau, 1993). Possibly, $ComF_B$ is an auxiliary protein, acquired after the branching of the *B. subtilis/amiloliquefaciens* group (Fig. 2a). It is interesting that a $comF_B$ homologue is also missing in the more distantly related *Streptococcus pneumoniae comF* operon (Berge *et al*., 2002).

The putative ComG-operons demonstrate some interesting features. Whereas $ComG_{AB}$ homologues (the ATPase and the polytopic membrane protein respectively) are present in all species analysed and $ComG_{CD}$ (similar to pilins) could also be confidently identified in at least the *B. cereus* linage (on the basis of genome location, homology and size of the encoded protein), no homologues of the $ComG_{EFG}$ proteins (minor pilins) were detected (*e*-value > E-1) in any other species than *B. subtilis*, *B. amyloliquefaciens*, *B. licheniformis* and *B. pumilus* (Fig. 1). In *B. subtilis*, all seven ORFs of the *comG* operon are required for efficient transformation (Chung and Dubnau, 1998), and it is conceivable that the absence of these genes explains the fact that no functional DNA uptake has been reported for these organisms. However, the genes downstream of $comG_D$ are conserved within the *B. cereus* group and it is plausible that they encode functional homologues of the *B. subtilis* $ComG_{EFG}$ proteins (Fig. 4). In spite of the lower homology of these proteins to *B. subtilis* $ComG_{EFG}$ at the level of whole protein, the N-terminal regions locally show a higher level of conservation (Fig. S2). The N-terminal region of minor pilins contains the conserved prepilin-like domain. Alternatively, $ComG_{ABCD}$ might be sufficient in these organisms. Notably, the $comG_A$ gene is truncated in *B. cereus* Cytotoxis strain and in contrast to the conserved DNA sequence within the *comG* operon no clear *orf* for $comG_E$ can be assigned, while the allotted *orf* for $comG_D$ is elongated (Fig. 4).

The *comG* operon of the alkalophilic *Bacilli* is disrupted by a tRNA island after $comG_B$. However, $ComG_{CD}$ proteins are encoded by genes downstream of the island, and at least in the case of *B. clausii* the genes downstream of $comG_D$ show strong homology to the *B. cereus* group of

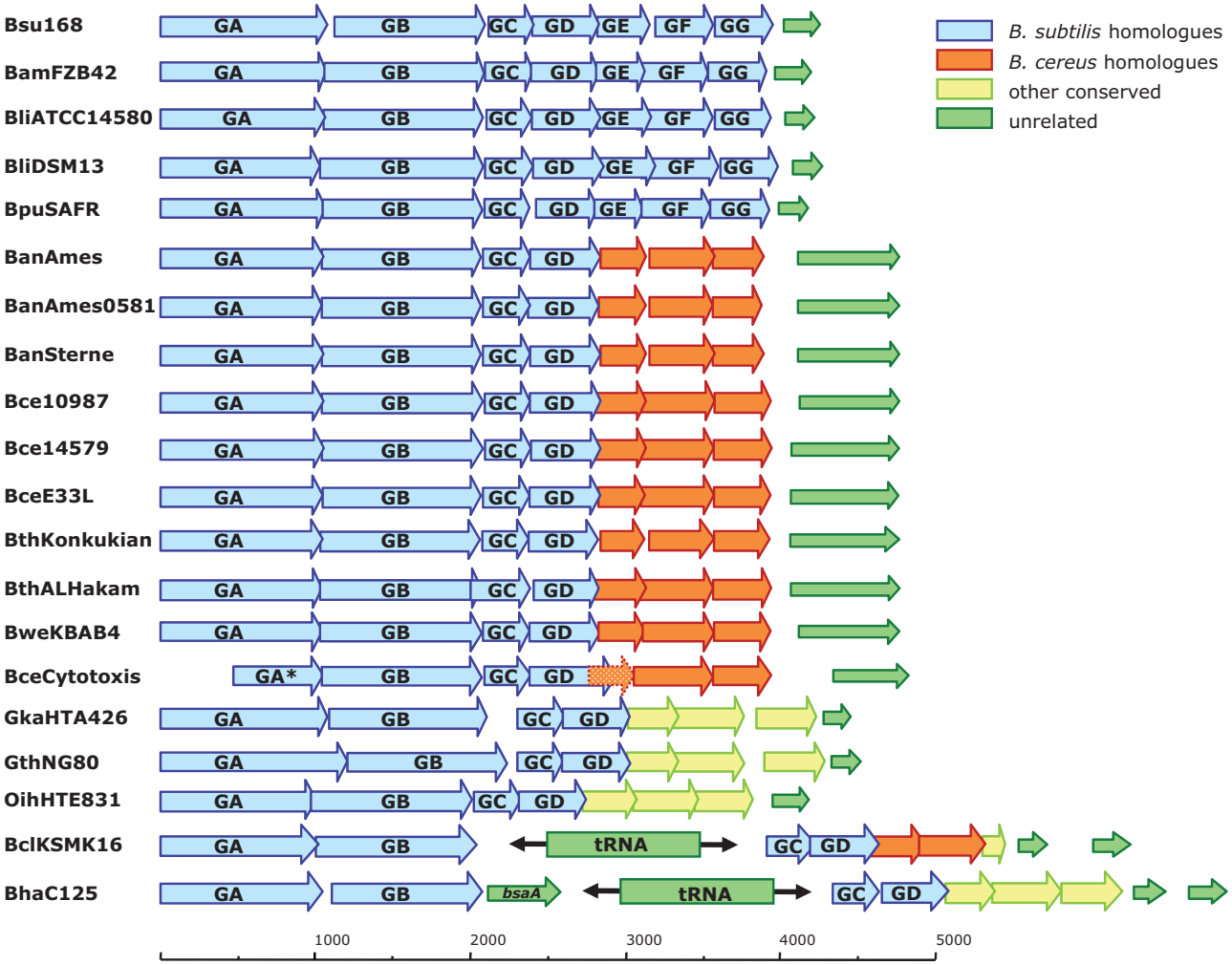

Fig. 4. The structure of the *comG* operon in *Bacillus* species. The putative functions of $ComG_A$, $ComG_B$, $ComG_C$ and $ComG_{DEFG}$ are traffic ATPase, polytopic membrane protein, major pilin and minor pilins respectively. For abbreviations of species names see Fig. 1. Genes in blue symbolize *B. subtilis comG* genes and homologues in other species. Orange is used for genes conserved in the *B. cereus* group and *B. clausii*, while yellow shows other conserved genes. Asterisk labels a truncated gene, while dotted arrow symbolizes region where no *orf* can be assigned, but high homology at DNA level.

downstream genes (BC4236-BC4235 genes in *B. cereus* ATCC14579 respectively) (Fig. 4). In fact, the genomic arrangement of genes homologue to the $comG_{ABCDEFG}$ is conserved even in *Streptococcus* species (*cglABCDEFG* in *S. pneumoniae*) (Peterson *et al*., 2004).

Induction of competence for DNA uptake in Gram-negative and Gram-positive bacteria

The conditions that trigger natural competence can vary greatly, and may include medium/nutrient limitation, growth phase, cell density and other stresses (Claverys *et al*., 2006). Additionally, it may depend on a given substrate in the environment, as was recently described for *Vibrio cholerae* (chitin) (Meibom *et al*., 2005) or occur only during a short period [e.g. *Streptococci* (Claverys *et al*., 2006) or *Acinetobacter baylyi* (Porstendorfer *et al*., 2000; Friedrich *et al*., 2001)]. As a result, establishing a competence regime can be a daunting task and this may in part explain the lack of successful transformation of species harbouring homologues of competence genes.

The mechanism of regulation of competence development generally diverges more than the DNA uptake machinery (Claverys and Martin, 2003), consistent with our findings in this study. Hence, several laboratories have focused on artificially inducing competence through the controlled expression of the putative key regulator. The ability to induce the expression of late competence genes has been shown for *Streptococcus pyogenes* (Woodbury *et al*., 2006), *Lactococcus lactis* (Wydau *et al*., 2006) and *Streptococcus thermophilus* (Blomqvist *et al*., 2006). However, only in the latter it was accompanied by

functional DNA uptake (Blomqvist *et al*., 2006), raising the question what prevents competence in the other instances. It has to be noted that in *Streptococci* the key regulator for the induction of genes for DNA uptake is a competence-specific sigma factor (ComX), different from the ComK protein described for *Bacilli* (Claverys *et al*., 2006).

Lack of transformation may be attributed to mutations or insertions that render essential competence proteins non-functional. For instance, the *comK* gene of *L. monocytogenes* is interrupted by the insertion of a prophage (Borezee *et al*., 2000). In this respect, the report that a clinical isolate of *S. pyogenes* seems to be capable of DNA transfer is noteworthy (Hidalgo-Grass *et al*., 2002). The complementation of putative non-competent strains with functional components of DNA-uptake machineries from closely related species may identify loss-of-function mutations.

Functional DNA uptake in various *Bacillus* species

In general, because of its dramatic impact on the physiology of the cells (Haijema *et al*., 2001), competence for genetic transformation commonly only develops under certain conditions – the so-called competence regime. In fact, *Neisseria gonorrhoeae* is the only organism reported to demonstrate constitutive competence (Sparling, 1966). Within the genus *Bacillus*, a competence regime has been established for *B. subtilis*, *B. licheniformis* and *B. amiloliquefaciens* (Spizizen, 1958; Thorne and Stull, 1966; Koumoutsi *et al*., 2004). There is, however, a great interest in the other members, from both a medical and biotechnological point of view. *Bacillus cereus* and *B. anthracis*, for instance, are causative agents of common food poisoning and anthrax respectively. *Bacillus halodurans* and *clausii* have both been isolated on the basis of their biotechnological potential. Artificially induced genetic competence, depending on the expression of a functional DNA uptake apparatus, can facilitate molecular studies on these bacteria.

Bacilli grow under diverse environmental conditions and group members are commonly observed in soil (Vilain *et al*., 2006), in the rhizosphere of a variety of plants (Fall *et al*., 2004), in food samples [*B. cereus* (Schoeni and Wong, 2005)], in insect guts [*B. thuringiensis* (Jensen *et al*., 2003)] or in an aquatic milieu (de Barros Soares *et al*., 2003). These natural habitats provide complex conditions that are not easily reproduced in the laboratory. It is therefore likely that the number of *Bacillus* species identified as competent is an underrepresentation of the actual occurrence of natural competence in this genus.

Recently, it has been demonstrated that *B. cereus* has the ability to take up DNA (Mironczuk *et al*., 2008). Upon the overproduction of *B. subtilis* ComK protein consistently low levels of transformation with both chromosomal or plasmid DNA were obtained. These results are significant for two reasons. First, it shows that the gene reservoir of *B. cereus* ATCC14579 is sufficient for the uptake and integration of DNA from the environment. This indicates either that the lack of homologues of the *B. subtilis* $comG_{EFG}$ is functionally complemented in *B. cereus* or that these are proteins not needed for functional DNA uptake in this organism. Second, as the observed levels of competence are much lower than obtained by the equivalent overexpression of ComK in *B. subtilis*, it is expected that other factors (regulatory and/or structural) will be able to augment efficient DNA uptake.

The regulation of transcriptional activation of late competence genes can be conserved between closely related species (Martin *et al*., 2006) and the artificial induction of the key regulator of competence in *S. thermophilus* resulted in DNA uptake (Blomqvist *et al*., 2006). We anticipate therefore that a strategy similar to Mirończuk *et al*. may induce functional DNA uptake in at least the close relatives of *B. cereus*, such as *B. anthracis* and *B. thuringiensis*. Moreover, the strategy may be applicable to induce competence in natural *Bacillus* isolates, some of which demonstrate traits such as the ability to form architecturally complex communities of cells (biofilm) that have been lost in laboratory strains (Branda *et al*., 2001).

Concluding remarks

By screening the genomes of fully sequenced *Bacillus* species, we have identified genes encoding homologues of the proteins involved in competence in *B. subtilis*. Our findings suggest that species for which no competence regime has been established so far have the potential to develop natural competence or acquire DNA. Indeed, *B. cereus* has recently been shown to harbour a functional DNA uptake machinery. The induction of DNA uptake, either naturally or induced, will facilitate molecular genetic studies with these organisms. Moreover, the insights gained from these comparative studies extend our understanding of natural competence in general and competence in *Bacillus* species specifically and gives directions for further research on the factors required for competence in the *Bacillus* genus.

Natural competence of *Bacilli* in the environment like in the rhizosphere of plants might contribute to the genome plasticity observed in *Bacilli*, similar to conjugation that was demonstrated to occur in the rhizosphere between *B. anthracis* species (Saile and Koehler, 2006).

Future studies on competence should be aimed at answering whether a single factor is capable of inducing late competence (analogous to ComK in *B. subtilis*) and what the genetic or environmental factors affecting the activity of these regulators are (e.g. by fusing the

promoter regions of ComK and/or late competence genes to reporter genes). Moreover, complementation studies will reveal whether putative (functional) homologues such as $ComG_{EFG}$ in the *B. cereus* group can support DNA uptake and may lead to the definition of a minimal competence machinery.

Acknowledgements

The authors wish to thank R.W.W. Brouwer for expert assistance with BLAST analysis and R. van Kranenburg for helpful discussions. A.T.K. was financially supported by PURAC and by Grant 818.02.004 from ALW-NWO Open programma, W.K.S. by Grant 811.35.002 and a Rubicon fellowship of the Netherlands Organization for Scientific Research. This project was carried out within the research programme of the Kluyver Centre for Genomics of Industrial Fermentation which is part of the Netherlands Genomics Initiative/ Netherlands Organization for Scientific Research.

References

Albano, M., Smits, W.K., Ho, L.T., Kraigher, B., Mandic-Mulec, I., Kuipers, O.P., and Dubnau, D. (2005) The Rok protein of *Bacillus subtilis* represses genes for cell surface and extracellular functions. *J Bacteriol* **187:** 2010–2019.

Anagnostopoulos, C., and Spizizen, J. (1961) Requirements for transformation in *Bacillus subtilis*. *J Bacteriol* **81:** 741–746.

Arnold, D.A., and Kowalczykowski, S.C. (2000) Facilitated loading of RecA protein is essential to recombination by RecBCD enzyme. *J Biol Chem* **275:** 12261–12265.

Averhoff, B. (2009) Shuffling genes around in hot environments: the unique DNA transporter of *Thermus thermophilus*. *FEMS Microbiol Rev* **33:** 611–626.

de Barros Soares, L.H., Assmann, F., and Zachia Ayub, M.A. (2003) Purification and properties of a transglutaminase produced by a *Bacillus circulans* strain isolated from the Amazon environment. *Biotechnol Appl Biochem* **37:** 295–299.

Berge, M., Moscoso, M., Prudhomme, M., Martin, B., and Claverys, J.P. (2002) Uptake of transforming DNA in Gram-positive bacteria: a view from *Streptococcus pneumoniae*. *Mol Microbiol* **45:** 411–421.

Berka, R.M., Hahn, J., Albano, M., Draskovic, I., Persuh, M., Cui, X., *et al.* (2002) Microarray analysis of the *Bacillus subtilis* K-state: genome-wide expression changes dependent on ComK. *Mol Microbiol* **43:** 1331–1345.

Blomqvist, T., Steinmoen, H., and Havarstein, L.S. (2006) Natural genetic transformation: a novel tool for efficient genetic engineering of the dairy bacterium *Streptococcus thermophilus*. *Appl Environ Microbiol* **72:** 6751–6756.

Borezee, E., Msadek, T., Durant, L., and Berche, P. (2000) Identification in *Listeria monocytogenes* of MecA, a homologue of the *Bacillus subtilis* competence regulatory protein. *J Bacteriol* **182:** 5931–5934.

Branda, S.S., Gonzalez-Pastor, J.E., Ben-Yehuda, S., Losick, R., and Kolter, R. (2001) Fruiting body formation by *Bacillus subtilis*. *Proc Natl Acad Sci USA* **98:** 11621–11626.

Chen, I., and Dubnau, D. (2004) DNA uptake during bacterial transformation. *Nat Rev Microbiol* **2:** 241–249.

Chen, I., Provvedi, R., and Dubnau, D. (2006) A macromolecular complex formed by a pilin-like protein in competent *Bacillus subtilis*. *J Biol Chem* **281:** 21720–21727.

Chung, Y.S., and Dubnau, D. (1995) ComC is required for the processing and translocation of *comGC*, a pilin-like competence protein of *Bacillus subtilis*. *Mol Microbiol* **15:** 543–551.

Chung, Y.S., and Dubnau, D. (1998) All seven comG open reading frames are required for DNA binding during transformation of competent *Bacillus subtilis*. *J Bacteriol* **180:** 41–45.

Claverys, J.P., and Martin, B. (2003) Bacterial 'competence' genes: signatures of active transformation, or only remnants? *Trends Microbiol* **11:** 161–165.

Claverys, J.P., Prudhomme, M., and Martin, B. (2006) Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annu Rev Microbiol* **60:** 451–475.

D'Souza, C., Nakano, M.M., and Zuber, P. (1994) Identification of *comS*, a gene of the *srfA* operon that regulates the establishment of genetic competence in *Bacillus subtilis*. *Proc Natl Acad Sci USA* **91:** 9397–9401.

Desai, B.V., and Morrison, D.A. (2006) An unstable competence-induced protein, CoiA, promotes processing of donor DNA after uptake during genetic transformation in *Streptococcus pneumoniae*. *J Bacteriol* **188:** 5177–5186.

Desai, B.V., and Morrison, D.A. (2007) Transformation in *Streptococcus pneumoniae*: formation of eclipse complex in a coiA mutant implicates CoiA in genetic recombination. *Mol Microbiol* **63:** 1107–1117.

Draskovic, I., and Dubnau, D. (2005) Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds. *Mol Microbiol* **55:** 881–896.

Dubnau, D. (1997) Binding and transport of transforming DNA by *Bacillus subtilis*: the role of type-IV pilin-like proteins – a review. *Gene* **192:** 191–198.

Dubnau, D. (1999) DNA uptake in bacteria. *Annu Rev Microbiol* **53:** 217–244.

Dubnau, D., and Cirigliano, C. (1972) Fate of transforming deoxyribonucleic acid after uptake by competent *Bacillus subtilis*: size and distribution of the integrated donor segments. *J Bacteriol* **111:** 488–494.

Eisenstadt, E., Lange, R., and Willecke, K. (1975) Competent *Bacillus subtilis* cultures synthesize a denatured DNA binding activity. *Proc Natl Acad Sci USA* **72:** 323–327.

Fall, R., Kinsinger, R.F., and Wheeler, K.A. (2004) A simple method to isolate biofilm-forming *Bacillus subtilis* and related species from plant roots. *Syst Appl Microbiol* **27:** 372–379.

Finkel, S.E., and Kolter, R. (2001) DNA as a nutrient: novel role for bacterial competence gene homologs. *J Bacteriol* **183:** 6288–6293.

Friedrich, A., Hartsch, T., and Averhoff, B. (2001) Natural transformation in mesophilic and thermophilic bacteria: identification and characterization of novel, closely related competence genes in *Acinetobacter* sp. strain BD413 and *Thermus thermophilus* HB27. *Appl Environ Microbiol* **67:** 3140–3148.

Hahn, J., Albano, M., and Dubnau, D. (1987) Isolation and characterization of Tn917lac-generated competence mutants of *Bacillus subtilis*. *J Bacteriol* **169:** 3104–3109.

Hahn, J., Maier, B., Haijema, B.J., Sheetz, M., and Dubnau, D. (2005) Transformation proteins and DNA uptake localize to the cell poles in *Bacillus subtilis*. *Cell* **122:** 59–71.

Haijema, B.J., Hamoen, L.W., Kooistra, J., Venema, G., and van-Sinderen, D. (1995) Expression of the ATP-dependent deoxyribonuclease of *Bacillus subtilis* is under competence-mediated control. *Mol Microbiol* **15:** 203–211.

Haijema, B.J., Hahn, J., Haynes, J., and Dubnau, D. (2001) A ComGA-dependent checkpoint limits growth during the escape from competence. *Mol Microbiol* **40:** 52–64.

Hamoen, L.W., Eshuis, H., Jongbloed, J., Venema, G., and van Sinderen, D. (1995) A small gene, designated *comS*, located within the coding region of the fourth amino acid-activation domain of *srfA*, is required for competence development in *Bacillus subtilis*. *Mol Microbiol* **15:** 55–63.

Hamoen, L.W., Van-Werkhoven, A.F., Bijlsma, J.J., Dubnau, D., and Venema, G. (1998) The competence transcription factor of *Bacillus subtilis* recognizes short A/T-rich sequences arranged in a unique, flexible pattern along the DNA helix. *Genes Dev* **12:** 1539–1550.

Hamoen, L.W., de Smits, W.K.J.A., Holsappel, S., and Kuipers, O.P. (2002) Improving the predictive value of the competence transcription factor (ComK) binding site in *Bacillus subtilis* using a genomic approach. *Nucleic Acids Res* **30:** 5517–5528.

Hamoen, L.W., Venema, G., and Kuipers, O.P. (2003) Controlling competence in *Bacillus subtilis*: shared use of regulators. *Microbiology* **149:** 9–17.

Han, C.S., Xie, G., Challacombe, J.F., Altherr, M.R., Bhotika, S.S., Brown, N., *et al.* (2006) Pathogenomic sequence analysis of *Bacillus cereus* and *Bacillus thuringiensis* isolates closely related to *Bacillus anthracis*. *J Bacteriol* **188:** 3382–3390.

Hidalgo-Grass, C., Ravins, M., Dan-Goor, M., Jaffe, J., Moses, A.E., and Hanski, E. (2002) A locus of group A *Streptococcus* involved in invasive disease and DNA transfer. *Mol Microbiol* **46:** 87–99.

Hoch, J.A. (1971) Genetic analysis of pleiotropic negative sporulation mutants in *Bacillus subtilis*. *J Bacteriol* **105:** 896–901.

Inamine, G.S., and Dubnau, D. (1995) ComEA, a *Bacillus subtilis* integral membrane protein required for genetic transformation, is needed for both DNA binding and transport. *J Bacteriol* **177:** 3045–3051.

Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V., *et al.* (2003) Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature* **423:** 87–91.

Jensen, G.B., Hansen, B.M., Eilenberg, J., and Mahillon, J. (2003) The hidden lifestyles of *Bacillus cereus* and relatives. *Environ Microbiol* **5:** 631–640.

Kidane, D., and Graumann, P.L. (2005) Intracellular protein and DNA dynamics in competent *Bacillus subtilis* cells. *Cell* **122:** 73–84.

Koumoutsi, A., Chen, X.H., Henne, A., Liesegang, H., Hitzeroth, G., Franke, P., *et al.* (2004) Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus amyloliquefaciens* strain FZB42. *J Bacteriol* **186:** 1084–1096.

Kramer, N., Hahn, J., and Dubnau, D. (2007) Multiple interactions among the competence proteins of *Bacillus subtilis*. *Mol Microbiol* **65:** 454–464.

Lindner, C., van Nijland, R.H.M., Bron, S., Hamoen, L.W., and Kuipers, O.P. (2004) Differential expression of two paralogous genes of *Bacillus subtilis* encoding single-stranded DNA binding protein. *J Bacteriol* **186:** 1097–1105.

Londono-Vallejo, J.A., and Dubnau, D. (1993) *comF*, a *Bacillus subtilis* late competence locus, encodes a protein similar to ATP-dependent RNA/DNA helicases. *Mol Microbiol* **9:** 119–131.

Londono-Vallejo, J.A., and Dubnau, D. (1994) Membrane association and role in DNA uptake of the *Bacillus subtilis* PriA analogue ComF1. *Mol Microbiol* **13:** 197–205.

Lorenz, M.G., and Wackernagel, W. (1994) Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev* **58:** 563–602.

Lovett, C.M., Love, P.E., and Yasbin, R.E. (1989) Competence-specific induction of the *Bacillus subtilis* RecA protein analog: evidence for dual regulation of a recombination protein. *J Bacteriol* **171:** 2318–2322.

Maier, B., Chen, I., Dubnau, D., and Sheetz, M.P. (2004) DNA transport into *Bacillus subtilis* requires proton motive force to generate large molecular forces. *Nat Struct Mol Biol* **11:** 643–649.

Martin, B., Quentin, Y., Fichant, G., and Claverys, J.P. (2006) Independent evolution of competence regulatory cascades in *Streptococci*? *Trends Microbiol* **14:** 339–345.

Meibom, K.L., Blokesch, M., Dolganov, N.A., Wu, C.Y., and Schoolnik, G.K. (2005) Chitin induces natural competence in *Vibrio cholerae*. *Science* **310:** 1824–1827.

Meima, R., Eschevins, C., Fillinger, S., Bolhuis, A., Hamoen, L.W., Dorenbos, R., *et al.* (2002) The *bdbDC* operon of *Bacillus subtilis* encodes thiol-disulphide oxidoreductases required for competence development. *J Biol Chem* **277:** 6994–7001.

Mironczuk, A.M., Kovacs, A.T., and Kuipers, O.P. (2008) Induction of natural competence in *Bacillus cereus* ATCC14579. *Microbial Biotech* **1:** 226–235.

Morrison, D.A., Mortier Barriere, I., Attaiech, L., and Claverys, J.P. (2007) Identification of the major protein component of the pneumococcal eclipse complex. *J Bacteriol* **189:** 6497–6500.

Mortier Barriere, I., Velten, M., Dupaigne, P., Mirouze, N., Pietrement, O., McGovern, S., *et al.* (2007) A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. *Cell* **130:** 824–836.

Ogura, M., Yamaguchi, H., Kobayashi, K., Ogasawara, N., Fujita, Y., and Tanaka, T. (2002) Whole-genome analysis of genes regulated by the *Bacillus subtilis* competence transcription factor ComK. *J Bacteriol* **184:** 2344–2351.

Palchevskiy, V., and Finkel, S.E. (2006) *Escherichia coli* competence gene homologs are essential for competitive fitness and the use of DNA as a nutrient. *J Bacteriol* **188:** 3902–3910.

Peterson, S.N., Sung, C.K., Cline, R., Desai, B.V., Snesrud, E.C., Luo, P., *et al.* (2004) Identification of competence pheromone responsive genes in *Streptococcus pneumo-*

niae by use of DNA microarrays. *Mol Microbiol* **51:** 1051–1070.
Piggot, P.J., and Coote, J.G. (1976) Genetic aspects of bacterial endospore formation. *Bacteriol Rev* **40:** 908–962.
Porstendorfer, D., Gohl, O., Mayer, F., and Averhoff, B. (2000) ComP, a pilin-like protein essential for natural competence in *Acinetobacter sp.* Strain BD413: regulation, modification, and cellular localization. *J Bacteriol* **182:** 3673–3680.
Prepiak, P., and Dubnau, D. (2007) A peptide signal for adapter protein-mediated degradation by the AAA+ protease ClpCP. *Mol Cell* **26:** 639–647.
Provvedi, R., and Dubnau, D. (1999) ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. *Mol Microbiol* **31:** 271–280.
Provvedi, R., Chen, I., and Dubnau, D. (2001) NucA is required for DNA cleavage during transformation of *Bacillus subtilis*. *Mol Microbiol* **40:** 634–644.
Rasko, D.A., Ravel, J., Okstad, O.A., Helgason, E., Cer, R.Z., Jiang, L., *et al.* (2004) The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *Bacillus anthracis* pXO1. *Nucleic Acids Res* **32:** 977–988.
Read, T.D., Peterson, S.N., Tourasse, N., Baillie, L.W., Paulsen, I.T., Nelson, K.E., *et al.* (2003) The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* **423:** 81–86.
Rey, M.W., Ramaiya, P., Nelson, B.A., Brody-Karpin, S.D., Zaretsky, E.J., Tang, M., *et al.* (2004) Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. *Genome Biol* **5:** R77.
Saile, E., and Koehler, T.M. (2006) *Bacillus anthracis* multiplication, persistence, and genetic exchange in the rhizosphere of grass plants. *Appl Environ Microbiol* **72:** 3168–3174.
Schoeni, J.L., and Wong, A.C. (2005) *Bacillus cereus* food poisoning and its toxins. *J Food Prot* **68:** 636–648.
van Sinderen, D., and Venema, G. (1994) *comK* acts as an autoregulatory control switch in the signal transduction route to competence in *Bacillus subtilis*. *J Bacteriol* **176:** 5762–5770.
van Sinderen, D., Kiewiet, R., and Venema, G. (1995) Differential expression of two closely related deoxyribonuclease genes, *nucA* and *nucB*. *Bacillus subtilis*. *Mol Microbiol* **15:** 213–223.
Smits, W.K., Hoa, T.T., Hamoen, L.W., Kuipers, O.P., and Dubnau, D. (2007) Antirepression as a second mechanism of transcriptional activation by a minor groove binding protein. *Mol Microbiol* **64:** 368–381.
Sonenshein, A.L. (2002) Developmental biology: regulation by selective gene localization. *Curr Biol* **12:** R90–R92.
Sparling, P.F. (1966) Genetic transformation of *Neisseria gonorrhoeae* to streptomycin resistance. *J Bacteriol* **92:** 1364–1371.
Spizizen, J. (1958) Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonuclate. *Proc Natl Acad Sci USA* **44:** 1072–1078.
Sturn, A., Quackenbush, J., and Trajanoski, Z. (2002) Genesis: cluster analysis of microarray data. *Bioinformatics* **18:** 207–208.
Susanna, K.A., Fusetti, F., Thunnissen, A.M., Hamoen, L.W., and Kuipers, O.P. (2006) Functional analysis of the competence transcription factor ComK of *Bacillus subtilis* by characterization of truncation variants. *Microbiology* **152:** 473–483.
Tadesse, S., and Graumann, P.L. (2007) DprA/Smf protein localizes at the DNA uptake machinery in competent *Bacillus subtilis* cells. *BMC Microbiol* **7:** 105.
Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., *et al.* (2000) Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res* **28:** 4317–4331.
Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22:** 4673–4680.
Thorne, C.B., and Stull, H.B. (1966) Factors affecting transformation of *Bacillus licheniformis*. *J Bacteriol* **91:** 1012–1020.
Turgay, K., Hamoen, L.W., Venema, G., and Dubnau, D. (1997) Biochemical characterization of a molecular switch involving the heat shock protein ClpC, which controls the activity of ComK, the competence transcription factor of *Bacillus subtilis*. *Genes Dev* **11:** 119–128.
Turgay, K., Hahn, J., Burghoorn, J., and Dubnau, D. (1998) Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. *EMBO J* **17:** 6730–6738.
Van de Peer, Y., and De Wachter, R. (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput Appl Biosci* **10:** 569–570.
Venema, G., Pritchard, R.H., and Venema-Schroder, T. (1965) Fate of transforming deoxyribonucleic acid in *Bacillus subtilis*. *J Bacteriol* **89:** 1250–1255.
Vilain, S., Luo, Y., Hildreth, M.B., and Brozel, V.S. (2006) Analysis of the life cycle of the soil saprophyte *Bacillus cereus* in liquid soil extract and in soil. *Appl Environ Microbiol* **72:** 4970–4977.
Woodbury, R.L., Wang, X., and Moran, C.P., Jr (2006) Sigma X induces competence gene expression in *Streptococcus pyogenes*. *Res Microbiol* **157:** 851–856.
Wydau, S., Dervyn, R., Anba, J., Dusko, E.S., and Maguin, E. (2006) Conservation of key elements of natural competence in *Lactococcus lactis* ssp. *FEMS Microbiol Lett* **257:** 32–42.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Presence of homologues of proteins involved in the regulation of competence in *Bacillus* and closely related species. Results of BLAST queries were visualized with Genesis 1.6 software: white is absent (with *e*-value of E-0), dark blue is present (*e*-value < E-20). BLAST analysis was performed with *B. subtilis* protein sequences against the translated protein database of a given genome. Question

marks denote small ORFs where identification is uncertain using the available bioinformatic tools that can miss homologues. Where the search yielded no hit (or with an *e*-value below E-05) an additional TBLASTN was performed. Protein names are indicated on the right. Bsu, *B. subtilis*; Bam, *B. amyloliquefaciens*; Bli, *B. licheniformis*; Bpu, *B. pumilus*; Ban, *B. anthracis*; Bce, *B. cereus*; Bth, *B. thuringiensis*; Bwe, *B. weihenstephanensis*; Oih, *O. iheyensis*; Gka, *G. kaustophilus*; Gth, *G. thermodenitrificans*; Bcl, *B. clausii*; Bha, *B. halodurans*).

Fig. S2. Multiple alignments of $ComG_{A-G}$ homologues. Black background represents conserved amino acids and grey background represents similar amino acids. Alignment was performed using Clustal W (Thompson *et al*., 1994), and presented using the Boxshade 3.21 program. For abbreviations of species names see Fig. S1. Conserved domains of *B. subtilis* ComG proteins are indicated with arrows above the alignment. AAA+, ATPase domain (Smart accession number: SM00382); GSPII-F, general secretion pathway domain (PFAM accession number: PF00482); TMS, transmembrane segment; N-met, N-methyl domain often found at N-terminus of pilins and proteins involved in secretion (PFAM accession number: PF07963); SSeq, signal peptide sequence.